A Variational Bayesian inference method for parametric imaging of PET data

M. Castellaro, G. Rizzo, M. Tonietto, M. Veronese, F.E. Turkheimer, M.A. Chappell, A. Bertoldo

PII: S1053-8119(17)30114-3
DOI: http://dx.doi.org/10.1016/j.neuroimage.2017.02.009
Reference: YNIMG13794

To appear in: *NeuroImage*

Received date: 13 October 2016
Revised date: 22 January 2017
Accepted date: 4 February 2017

Cite this article as: M. Castellaro, G. Rizzo, M. Tonietto, M. Veronese, F.E. Turkheimer, M.A. Chappell and A. Bertoldo, A Variational Bayesian inference method for parametric imaging of PET data, *NeuroImage* http://dx.doi.org/10.1016/j.neuroimage.2017.02.009

A Variational Bayesian inference method for parametric imaging of PET data

Castellaro M.[1], Rizzo G.[1], Tonietto M.[1], Veronese M.[2], Turkheimer F.E.[2], Chappell M.A.[3], Bertoldo A.[1]

[1] Department of Information Engineering, University of Padova, Italy

[2] Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

[3] Institute of Biomedical Engineering, University of Oxford, Old Road Campus Research Building, Old Road Campus, Headington, Oxford, United Kingdom

Part of this work was presented at the World Molecular Imaging Congress (Honolulu, September 2015)

Corresponding author:

Alessandra Bertoldo

Department of Information Engineering (DEI)
University of Padova
Via G. Gradenigo 6/B
35131 Padova
Italy
Phone: (+39)-049-8277694
Fax: (+39)-049-8277699
E-mail bertoldo@dei.unipd.it

Running headline: Variational Bayesian inference for PET quantification

Abstract

In dynamic Positron Emission Tomography (PET) studies, compartmental models provide the richest information on the tracer kinetics of the tissue. Inverting such models at the voxel level is however quite challenging due to the low signal-to-noise ratio of the time activity curves. In this study, we propose the use of a Variational Bayesian (VB) approach to efficiently solve this issue and thus obtain robust quantitative parametric maps.

VB was adapted to the non-uniform noise distribution of PET data. Moreover, we propose a novel hierarchical scheme to define the model parameter priors directly from the images in case such information are not available from the literature, as often happens with new PET tracers.

VB was initially tested on synthetic data generated using compartmental models of increasing complexity, providing accurate (%bias<2%±2%, root mean square error<15%±5%) parameter estimates. When applied to real data on a paradigmatic set of PET tracers (L-[1-^{11}C]leucine, [^{11}C]WAY100635 and [^{18}F]FDG), VB was able to generate reliable parametric maps even in presence of high noise in the data (%unreliable estimates<11%±5%).

Highlights

- Variational Bayesian (VB) approach is applied for the first time to PET data.
- VB was adapted to the specific non-uniform noise distribution of PET data.
- Various PET tracers described by different compartmental models were tested.
- VB provided robust and accurate model estimates with low percentage of unreliable estimates.

Keywords

Positron Emission Tomography, Variational Bayes, Voxel-wise kinetic analysis, L[1-^{11}C]leucine, [^{18}F]FDG, [^{11}C]WAY-100635

1. Introduction

Quantitative brain imaging with dynamic Positron Emission Tomography (PET) based on compartmental models can be performed at both region of interest (ROI) or voxel level. The analysis at the ROI level offers the advantage of a higher signal-to-noise ratio (SNR) allowing a more accurate and precise numerical identification of the model parameters. On the contrary, ROI analysis is prone to the intrinsic loss of the original spatial resolution and to the tacit assumption that the between-voxel variability of the time activity curves (TACs), averaged within the ROI, can be ignored or is not significant for the results. Voxel-wise quantification based on the full kinetic modelling overcomes these limitations but, in turn, it is hampered by the low SNR of the TACs derived from the single voxel. Nonlinear least squares estimators, which are considered the gold standard for ROI-based quantification, are too sensitive to the noise in the data at the voxel level and therefore parameter estimates are characterized by either high percentage of non-physiological estimates, lack of convergence or low precision (i.e. coefficient of variation of estimated parameters greater than 100% (DiStefano, 2015), page 546). Thus, the development of reliable and general-purpose parametric imaging methods remains a challenge for dynamic quantitative PET imaging. Nevertheless, during the last decade, several approaches have been proposed. Among them, the basis function method (Gunn et al., 1997; Koeppe et al., 1985; Rizzo et al., 2013a; Tomasi et al., 2009) is by far the most used for parametric imaging based on compartmental modelling, but its applicability is restricted to simple compartmental model structures. Other classical, albeit simplified, solutions for parametric imaging in PET are represented by graphical approaches (Logan et al., 1990; Patlak et al., 1983) and Spectral Analysis methods (Cunningham and Jones, 1993; Turkheimer et al., 2003; Veronese et al., 2012) but these approaches do not fully solve the underlying compartmental model (e.g. Spectral Analysis)

and they often do not return any information on the micro-parameters (individual compartmental rate constants of the model), as is the case for graphical methods. The identification of such micro-parameters is useful to fully characterize the physiology of the system, since there can be changes in pathological states which are not only linked to macro-parameters of interest (Kotasidis et al., 2014): for example, a recent study found that skeletal muscle insulin resistance in type-2 diabetes involves a severe impairment of glucose transport and additional impairment in the efficiency of glucose phosphorylation (Goodpaster et al., 2014). In addition, the use of micro-parameters estimation at the voxel level has been shown to permit the identification of a pattern of cholinergic dysfunction in Alzheimer Disease (Marcone et al., 2012). Moreover, in the oncology field, fully quantitative parameters based on kinetic modelling could complement or even supersede semi-quantitative analysis in the clinical practice (Kotasidis et al., 2014).

A valid alternative for parametric mapping is represented by Bayesian methods, which incorporate prior information on the tissue kinetics and have already been adopted in the PET community (Alpert and Yuan, 2009; Peng et al., 2008; Rizzo et al., 2012; Zanderigo et al., 2010; Zhou et al., 2013). Prior information could also be included in hybrid approaches that combine reconstruction and kinetic modelling (Kamasak et al., 2005). Moreover, it is possible to incorporate structural and anatomical information obtained from magnetic resonance imaging (Loeb et al., 2015; Tang et al., 2010).

However, despite providing robust estimates also at the voxel level, Bayesian methods proposed so far often do not have the required flexibility to be generalized to at least all the most common compartmental models, and when they do, their applicability to the clinical practice is undermined by the high computational time required for analysing a whole brain dynamic PET scan (up to several hours). For example, the Bayesian methods proposed by

Alpert and Yuan (2009) and Zanderigo et al. (2010) are based on nonlinear estimators and require high computational time. Furthermore, in (Alpert and Yuan, 2009), the prior information are obtained by analyzing a prior cohort of parametric images and there is therefore the necessity of having an additional sufficiently large data set to derive reliable a priori information. In Zhou et al. (2013), the quantification problem is tackled in a full Bayesian framework, solving the model with a Markov chain Monte Carlo (MCMC) sampling approach. In Peng et al. (2008) and Rizzo et al. (2012) the compartmental model is solved at the voxel level by linearizing the model (with an over-complete exponential basis set in Peng et al. (2008)) and then solving it using a Sparse Bayesian Learning or Maximum A Posteriori approach, respectively.

The critical point of any Bayesian approach is the computation of the posterior distribution derived by the Bayes' rule. Unfortunately, the numerical integrations involved are often computationally intractable. Sampling approaches, for example MCMC, are generally employed to calculate a numerical approximation of the posterior distribution. However, while these approaches are asymptotically exact, they are still too computationally expensive to be used for nonlinear Bayesian inference at the voxel level. An alternative is the Variational Bayesian (VB) method (Chappell et al., 2009), which is a fully Bayesian approach that uses an analytical approximation to simplify the calculation of the posterior distribution; this approximation is numerically tractable for both linear and nonlinear systems, including linear, time-invariant compartmental models. Whilst VB has previously been applied in the PET field for reconstruction and segmentation (Rapisarda et al., 2014; Xia et al., 2011), it has never been applied for kinetic modelling.

In the current study, VB is adapted and applied for quantitative parametric mapping of PET data. In order to demonstrate that VB is suitable as general Bayesian framework for

quantitative dynamic PET data, we first customised the VB algorithm to the peculiarities of the noise distribution in PET kinetic data (Zanoni et al., 2015). Secondly, we assessed VB performance using synthetic data generated using compartmental models having varying complexity, and, finally proposing a novel data driven prior generation, we applied VB to real data on a paradigmatic set of tracers representative of the variety of models used for PET quantification.

2. Material and methods

2.1. Theoretical framework of Variational Bayes

In a Bayesian parameters estimation approach, *a priori* information is used to aid the numerical identification of the vector of the parameters $\boldsymbol{\theta}$ of a chosen model w from a set of measured data $\boldsymbol{y}$. Bayes' theorem links the calculation of the posterior distribution of the parameters given the data and the model ($P(\boldsymbol{\theta}|\boldsymbol{y},w)$) to the *a priori* distributions of the parameters to be estimated ($P(\boldsymbol{\theta}|w)$). This is obtained through the likelihood ($P(\boldsymbol{y}|\boldsymbol{\theta},w)$), the probability density function that describes the data given the parameters and the model. Bayes' rule can be written in a simpler form neglecting the dependence on the chosen model w as:

$$P(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\boldsymbol{y})} \quad (1)$$

In real applications, the numerical integrations needed for the direct computation of the posterior are usually intractable. Variational Bayesian approaches analytically approximate the actual posterior with a simpler form $Q(\boldsymbol{\theta})$ (Attias, 2000), with the computational burden shifted to maximizing the agreement between true and approximate posterior. The distance between the approximation $Q(\boldsymbol{\theta})$ and the true posterior distribution of the parameters $P(\boldsymbol{\theta}|\boldsymbol{y})$

can be measured via the Kullback–Leibler divergence $\mathrm{KL}[Q(\boldsymbol{\Theta})||P(\boldsymbol{\Theta}|\boldsymbol{y})]$ which, however, cannot be computed without knowledge of the true posterior. Nonetheless, KL is equal to (Beal, 2003):

$$\mathrm{KL}[Q(\boldsymbol{\Theta})||P(\boldsymbol{\Theta}|\boldsymbol{y})] = \log P(\boldsymbol{y}) - \mathrm{F} \tag{2}$$

where F is the Free Energy which is defined as:

$$\mathrm{F} = \int Q(\boldsymbol{\Theta}) \log \frac{P(\boldsymbol{y}|\boldsymbol{\Theta})P(\boldsymbol{\Theta})}{Q(\boldsymbol{\Theta})} \, \mathrm{d}\boldsymbol{\Theta} \tag{3}$$

and since $\log P(\boldsymbol{y})$ does not depend on $\boldsymbol{\Theta}$ and the Kullback–Leibler divergence is always non-negative, this latter can be minimized by maximizing F.

To make the integrals tractable VB specifies a mean field approximation for $Q(\boldsymbol{\Theta})$. This consists in collecting the vector of parameters $\boldsymbol{\Theta}$ into separate groups. In this work, one group included all the model parameters ($\boldsymbol{\theta}$) while the other only the parameter controlling for noise precision (ϕ). Each group is described by its own approximate posterior distribution ($Q_\theta(\boldsymbol{\theta}|\boldsymbol{y})$ and $Q_\phi(\phi|\boldsymbol{y})$), which are assumed independent between them (i.e. $Q(\boldsymbol{\Theta}) = Q_\theta(\boldsymbol{\theta}|\boldsymbol{y}) \cdot Q_\phi(\phi|\boldsymbol{y})$). Furthermore, the use of prior conjugated with the likelihood, i.e. with the same parametric form of the posterior, simplifies the computation of the factorized posteriors, as the VB update becomes a process of updating the posterior hyper-parameters. The interested reader is referred to the original publications for a detailed derivation of the method (Attias, 2000; Chappell et al., 2009).

The method section is organized as follows: first, we will present the modifications required to apply VB to PET data, with a particular focus on the extension of the error model to non-uniform noise and the novel data-driven derivation of the priors. Then, we will present

several scenarios of simulated and clinical PET data that represent an extended set of case studies for the application of VB.

2.2. Adapting the noise model to PET applications

The model for the $N \times 1$ vector of PET measurements $\boldsymbol{y}(\boldsymbol{t})$ (representing the tracer concentration over time for a given voxel) is:

$$\boldsymbol{y}(\boldsymbol{t}) = g(\boldsymbol{\theta}, \boldsymbol{t}) + \boldsymbol{e}(\boldsymbol{t}) \tag{4}$$

where $g(\boldsymbol{\theta}, \boldsymbol{t})$ is the compartmental model with the parameter vector $\boldsymbol{\theta}$ (which represents all the individual rate constants of the compartmental model) and $\boldsymbol{e}(\boldsymbol{t})$ is additive Gaussian noise. For the sake of clarity, the dependency from the time t has been omitted in the following passages.

In its original version, the VB approach assumed a constant level of noise for all the measurements, with zero mean and variance equal to ϕ^{-1}, i.e. $\boldsymbol{e} \sim N(0, \phi^{-1}\mathbf{I})$. The non-uniform sampling grid used in PET applications implies that, during each measured sample, the scintillation crystals are open to record the coincident pairs of photons emitted by the tracer for different time durations. The length of the time frames is chosen in order to detect a sufficient numbers of decay events to reconstruct the image. Hence, the level of noise will be different at each time point. This can be incorporated in the VB framework by introducing a non-identity covariance matrix $\boldsymbol{\Sigma_e}$ in the noise model:

$$\boldsymbol{e} \sim N(0, \phi^{-1}\boldsymbol{\Sigma_e}) \tag{5}$$

which is a diagonal matrix whose elements are calculated as the ratio between the measured tracer concentration $\boldsymbol{y}$ and the duration of the time frame $\Delta\boldsymbol{t}$, i.e. for the i-th time point $\Sigma_e(t_i, t_i) = y(t_i)/\Delta t_i$, as standard practice in dynamic PET studies (Bertoldo et al., 1998; Mazoyer et al., 1986). Nevertheless, in the first frames, this formula tends to underestimate

the noise relative variance since the tracer has not yet reached the voxel and therefore the term $y(t_i)$ is very small. This will be reflected in an exaggerated influence of the correspondent samples in the fitting procedure, providing an artificial overfitting. Therefore, in order to restrict their contribution on the likelihood (Zhou et al., 2013), we truncated the first time points' precisions (assuring a maximum possible ratio of 10 between all precisions). The full modified equations for the VB update and the Free Energy formulation are reported in Appendix A.

2.3. Data-driven prior definition

The prior distributions chosen in this work are a multivariate normal distribution (MVN) for the vector of model parameters $\boldsymbol{\theta}$ and a Gamma distribution for the noise parameter ϕ:

$$P(\boldsymbol{\theta}) \sim \mathrm{MVN}(\boldsymbol{\theta}; \boldsymbol{m_0}, \boldsymbol{\Lambda_0^{-1}}) \tag{6}$$

$$P(\phi) \sim \mathrm{Ga}(\phi; s_0, \mathrm{c}_0) \tag{7}$$

and the factorized posteriors are chosen conjugate with the priors:

$$Q_\theta(\boldsymbol{\theta}|\boldsymbol{y}) \sim \mathrm{MVN}(\boldsymbol{\theta}; \boldsymbol{m}, \boldsymbol{\Lambda^{-1}}) \tag{8}$$

$$Q_\phi(\phi|\boldsymbol{y}) \sim \mathrm{Ga}(\phi; s, \mathrm{c}) \tag{9}$$

where $\boldsymbol{m}$ and $\boldsymbol{\Lambda^{-1}}$ are the mean vector and the covariance matrix of the MVN (with $\boldsymbol{m_0}$ and $\boldsymbol{\Lambda_0^{-1}}$ as corresponding prior mean and covariance) and s and c are the shape and scale parameter of the Gamma distribution that parametrizes the noise variance (with s_0 and c_0 as corresponding prior values).

The prior distribution of the parameters of the model can be set to literature values. However, it can be difficult to retrieve this kind of information, especially when the model is complex, or the tracer relatively new, or even when the quantification is performed on

patients, i.e. representing pathological conditions that can hamper the range of normality of the parameters. Nevertheless, it is possible to retrieve robust and reliable information directly from the data which can be used in place of strict prior information. Here we propose an approach similar to Empirical Bayesian methods (Casella, 1985), in which a weak prior distribution is estimated for the parameters using the data itself. In this scheme, the prior distribution is defined at the voxel level following a hierarchical scheme, where the estimates obtained from model-fitting at the region level with weighted nonlinear least squares are passed to the voxel layer as a priori information (WNLLS) (Rizzo et al., 2013a). WNLLS is considered the gold standard at the ROI level, and hence it is plausible that the prior retrieved is valuable.

Figure 1 reports the workflow of the prior definition in a hierarchical fashion. Regional TACs can be obtained either by anatomical segmentation or by functional clustering. The compartmental model is solved for each ROI TAC by using WNLLS and the regional estimates are used as the MVN prior mean ($\boldsymbol{m_0}$) of the model parameters for all the voxels composing the selected region. In this way, different regions employ a tailored prior, but variations in the parameters value at the voxel level are still permitted in the inference procedure.

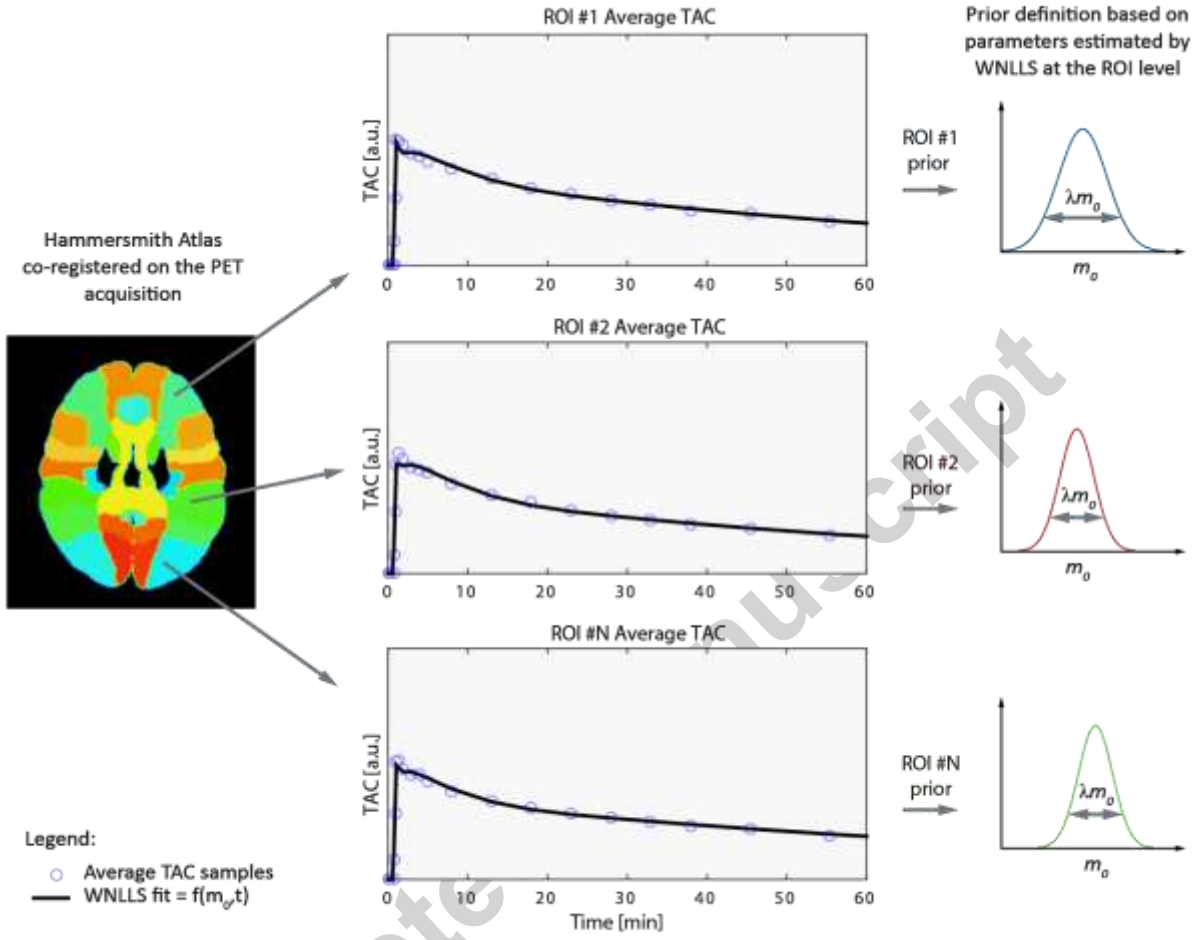

Figure 1. Workflow for the definition of the a priori distribution exploiting the hierarchical approach. *The PET acquisition is sub-divided in regions of interest (ROIs) based on an anatomical atlas segmentation or functional clustering. A WNLLS estimator is used on the average TAC of each ROI to retrieve the mean of each parameter's prior distribution (*$\boldsymbol{m_0}$*). For three representative ROIs the mean TAC (blue circles) and the fit obtained with the WNLLS estimator (black solid line) are reported. The prior distributions are defined from the WNLLS estimates and then hierarchically translated to each voxel of each*

ROI. The precisions of priors are obtained using the formula $\boldsymbol{\Lambda_0} = \frac{1}{(\lambda\, \boldsymbol{m_0})^2}$ *(Eq. 10) where* λ *is chosen by simulations (see Data-driven prior definition section).*

As regard the precision of the model prior ($\boldsymbol{\Lambda_0}$), this must be set based on the level of variability in the estimates: a low variance will anchor the posterior mean to that of the prior distribution and a high level of variance will allow the parameters to be freely estimated from the (noisy) voxel level data. In the extreme case of an uninformative prior this would reduce to the WNLLS solution. Since it is very difficult to retrieve robust a priori information on the co-variances, the off-diagonal elements of $\boldsymbol{\Lambda_0}$were set to zero. This limitation can introduce a modest structure in the relation between parameters, considering them in principle independents. This is however not a restriction, since the estimator can infer the co-variances in the estimation process when supported from the data.

We set the diagonal elements of $\boldsymbol{\Lambda_0}$ based on the region-wise WNLLS estimates (which correspond to the prior mean $\boldsymbol{m_0}$), multiplied by λ, which gives a measure of the expected variability across the brain as:

$$\boldsymbol{\Lambda_0^{i,j}} = \begin{cases} \dfrac{1}{\left(\lambda\, \boldsymbol{m_0^i}\right)^2} & i = j \\ 0 & otherwise \end{cases} \qquad (10)$$

Under the assumption that $\boldsymbol{m_0}$ coincides with the mean of the distribution of parameters inside the ROI, λ corresponds to its coefficient of variation across voxels. This cannot be derived directly from the average TAC obtained at the ROI level and its value must be chosen in advance. This is done in a simulated environment by varying λ on a suitable grid of values (from 5% to 200%), i.e. by varying the prior variance. The optimal λ was defined in a

simulated scenario, with a sensitivity analysis (see section Simulation 2), and then applied to real cases.

2.4. Application to simulated data

Simulation studies were performed to:

1) Assess the performance of VB compared to WNLLS in a paradigmatic sets of compartmental models: the two-tissue four-rate constants compartmental model (2TCM) simulated with fast (Rizzo et al., 2013a) and slow (Rizzo et al., 2013b) kinetics, and three-tissue five-rate constants compartmental model (Bertoldo's 5K model) (Bertoldo et al., 2001).

2) Derive a measure of the variability λ and evaluate the impact of a prior generated from the data at the ROI level on the VB estimates at the voxel level (sensitivity study).

2.5. Simulation study 1: performance of VB and WNLLS

The simulation was set up to generate synthetic TACs characterized by a signal-to-noise ratio comparable to voxel-level activities. We focused our analysis on the 2TCM model (the most common model in neuroreceptor studies) and 5K model (the compartmental model for [^{18}F]FDG in skeletal muscle, a model more complex than the usual 2TCM). The model structure, equations and parameters of interest are reported in Figure 2A and 2C.

Details on the tracers and ROIs used to simulate the synthetic data are reported in Table 1. The main steps of the simulation are showed in a pseudo-code fashion and details required for its implementation are reported in the following.

Table 1.Tracers used to simulate the synthetic dataset. For each tracer the kinetics characteristic, the regions of interest considered and the reference to the original publication of the data are reported.

Tracer	Kinetics characteristics	Regions of interest	Reference
[^{11}C](R)-rolipram	Slow 2TCM kinetics	frontal cortex, thalamus and putamen	(Rizzo et al., 2013b)
[^{11}C]WAY100635	Fast 2TCM kinetics	cerebellum, frontal and temporal cortex	(Rizzo et al., 2013a)
[^{18}F]FDG	Irreversible 5K kinetics	soleus and tibialis anterior	(Bertoldo et al., 2001)

Pseudo-Code of Simulation 1: performance of VB and WNLLS
nT = number of tracers
nR = number of ROIs
nV = number of voxels
nSV = number of synthetic voxels
nMC = number of Monte Carlo realizations

for each tracer (t)
  for each ROIs (r)
    for each voxel (v)
    $\hat{p}$(v,r,t) = WNLLS_estimation(voxelTAC(v,r,t));
    end
  end
  $\hat{p}_{OK}$ = reliable and physiological estimates

  % definition of true values
  MVN_p_{true} = MVN(mean($\hat{p}_{OK}$), COV($\hat{p}_{OK}$));

  % prior definition
  m_0 = mean(MVN_p_{true});
  $\Lambda_0 = \left(3\ \text{std}(MVN\_p_{true})\right)^{-2}$;

for each synthetic voxel (sv)
  $TAC_{noise-free}$ = $model_t$(time, $MVN\_p_{true}(sv)$);

  for each Monte Carlo realization (mcr)
    TAC_{noisy} = $TAC_{noise-free}(:, sv)$ + noise(:,mcr);

$\hat{p}_{WNLLS}$(sv,mcr,t) = WNLLS_estimation(TAC_{noisy});
$\hat{p}_{VB}$(sv,mcr,t) = VB_estimation(TAC_{noisy}, m_0, Λ_0);
end *% Monte Carlo realization*
end *% synthetic voxel*
end *% tracer kinetics*

The VB estimation step can be summarized as:

MaxIter = 100
MaxTrials = 10
Tolerance = 10^{-6}

for i from 1 to MaxIter
Update noise, model parameters ($\hat{p}_{VB}$) (Eq. A1-4) and FreeEnergy (F_{new}) (Eq. 5)
if $F_{new} = F_{old} \pm$ tolerance
Return
else if $F_{new} > F_{old}$
Update noise, model parameters (Eq. A1-4) and FreeEnergy (Eq. 5)
Trials = 0;
else if ($F_{new} < F_{old}$) & (trials < Maxtrials)
Save current solution ($\hat{p}_{VB}$)
Update noise, model parameters (Eq. A1-4) and FreeEnergy (Eq. 5)
Increment Trials
else
Discard further updates and provide saved solution
Return
end *%termination criteria*
end %VB iteration

For each tracer, the reliable voxel-wise estimates in a subset of the regions of interest (ROI) (reported in Tab.1) were selected from a representative subject. The ROIs were selected to represent the major peculiarities of the tracer kinetics. Each voxel was estimated using a WNLLS estimator. Then, the micro-parameter estimates, from every voxel of all the ROIs, were used to calculate the sample mean and covariance with the aim to define a MVN distribution representative of each tracer kinetics. Then, the unreliable and not physiological estimates were excluded, according to two criteria: 1) at least one micro-parameter should show a coefficient of variation greater than 100%, 2) the estimator did not reach convergence (maximum number of iterations reached). The coefficient of variation represents the precision of the parameter estimates and is calculated as the ratio between the estimated standard

deviation (derived by the inverse of the Fisher-information matrix when using WNLLS and from the posterior distribution in the VB case) and the expected value of the parameter. If the CVs are too high (e.g. CV>100%), the model is not a posteriori or numerically identifiable and should be rejected (Cobelli et al., 2002; DiStefano, 2015).

To create a set of synthetic voxels comparable to an in-vivo dataset and in order to test both the VB and WNLLS performances, 1000 sets of micro-parameters were drawn from each MVN (i.e. we obtained 1000 sets of 5 parameters for both slow and fast 2TCM kinetics (K_1, k_2, k_3, k_4, V_b) and 1000 sets of 6 parameters for 5K model (K_1, k_2, k_3, k_4, k_5, V_b)). We also derived the macro-parameters of interest for the two models (V_T [ml/cm^3] volume of distribution for the 2TCM and K_i [ml/cm^3/min] irreversible uptake of the tracer in the tissues for 5K model).

From each set of parameter an impulsive response function was generated accordingly to equations reported in Figure 2A and Figure 2C. Then, a noise-free TAC for each set of micro-parameter, that represents a synthetic voxel, was generated by convolving the impulse response function with a metabolite-corrected arterial input function measured in previous studies.

Each noise-free synthetic voxel TAC was used in a Monte Carlo simulation with 1000 independent noise realizations. The noise variance was defined as in Eq. 5, where the proportionality constant ϕ^{-1} was fixed from the real data of the original studies (estimated a posteriori as in (Bertoldo et al., 1998)).

In summary, 3 million TACs (1000 Monte Carlo realizations x 1000 synthetic voxels x 3 tracer's kinetics) were simulated and they were quantified with both WNLLS and VB. The priors for the VB estimator were independently generated for each tracer kinetics defining $\boldsymbol{m_0}$

as the mean of the MVN distribution used to generate the synthetic voxels and $\boldsymbol{\Lambda_0}$ as the variability (as 3 SD) across the synthetic voxel set. These prior means were also used as initial values for WNLLS.

The weights used in WNLLS were set as the inverse of the variance of the PET measurement error whose definition was the same as for VB. The unknown scale factor ϕ^{-1} is intrinsically considered in the VB algorithm as $1/(s \cdot c)$ from the gamma distributed factorized posterior, while it was estimated a posteriori as in (Bertoldo et al., 1998) for WNLLS.

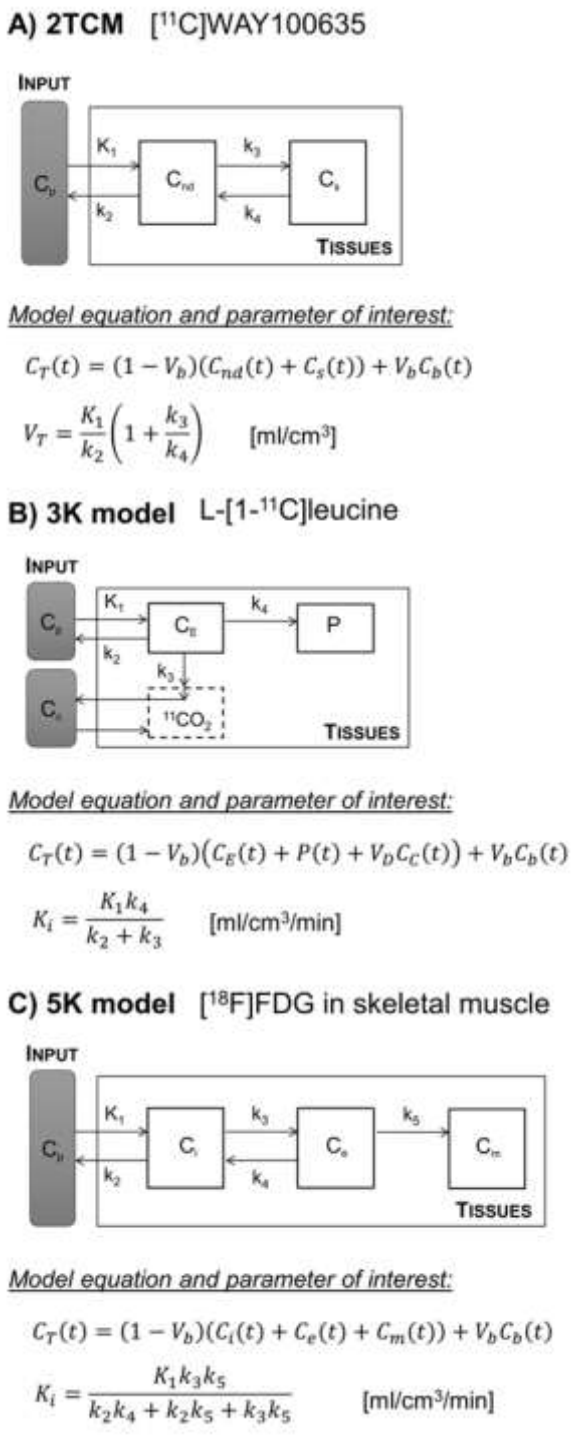

Figure 2. Summary of compartmental models and main outcomes The two-tissue four-rate constant compartmental model (2TCM, A), the two-tissue three-rate constants (3K model, B) and the three-tissue five-rate constants (5K model, C) are presented, along with the corresponding model equations and the main parameters of interest (V_T, volume of distribution, [ml/cm^3] and K_i, net uptake rate constant, [ml/cm^3/min]). Legend: K_1 [ml/cm^3/min], k_2 [1/min], k_3 [1/min], k_4 [1/min], k_5 [1/min], C_p, arterial plasma; C_{nd}, non displaceable binding; C_s, specific binding; C_c,

blood $^{11}CO_2$; C_E, free L-[1-^{11}C]leucine; P, L-[1-^{11}C]leucine incorporated into tissue protein; V_D, brain–blood equilibrium $^{11}CO_2$ distribution volume; C_i, intracellular concentration; C_e, extracellular concentration; C_m, metabolized [^{18}F]FDG-6-phosphate intracellular concentration.

2.6. Simulation study 2: prior sensitivity

The second simulation aimed to evaluate the sensitivity of VB to both prior mean and precision. The simulation can be employed as a method to determine the optimal value of λ, which represents a measure of variability of the parameters across the whole brain. As for simulation 1, the main steps of simulation 2 are reported for the sake of clarity in a pseudo-code fashion, whereas the details on the implementation are reported in the following.

Pseudo-Code of Simulation 2: sensitivity to the prior

for each tracer (t)
  clusters = kmeans(data(t));
  for each cluster with #voxels > 1000 (c)
    for each voxel (v)
      $\hat{p}$(v,c,t) = WNLLS_estimation(voxelTAC(v,c,t));
    end

    $\hat{p}_{OK}$ = reliable and physiological estimates

    % definition of true values
    MVN_p_{true} = MVN(mean($\hat{p}_{OK}$), COV($\hat{p}_{OK}$));
    % prior mean definition as real case
    m_0 = WNLLS_estimation(cluster(c).centroid);

    for each synthetic voxel (sv)
      $TAC_{noise-free}$ = $model_t$(time, $MVN\_p_{true}(sv)$);
      TAC_{noisy} = $TAC_{noise-free}(:, sv)$ + noise(sv);

```
        for each λ from 5% to 200% with step of 5%
            % prior precision definition
            Λ0 = (λ m0)^-2;
            p̂VB(sv,c,t) = VB_estimation(TAC_noisy, m0, Λ0);
        end % λ
    end % synthetic voxel
  end % cluster
end % tracer kinetics
```

For each tracer, the same representative subject used in simulation 1, selected from the available datasets, was analyzed with a K-means algorithm, with a fixed number of 20 clusters, 1000 iterations and 10 replicates. The clusters with less than 1000 voxels were excluded from further analysis. For the remaining clusters each voxel was analyzed with the WNLLS estimator. The micro-parameters estimated were then filtered with the same criteria used in simulation 1, including only physiological solution and reliable estimates. The remaining voxel estimates were used to estimate mean and co-variance of a MVN distribution, that was then used to sample 1000 noise-free synthetic voxels TAC following the same procedure as in simulation 1. Each noise-free synthetic voxels TAC was then independently corrupted with additive noise. The variance of the noise was set following Eq. 5. The prior mean ($\boldsymbol{m_0}$), to be used with VB, was defined from the mean TAC of each synthetic voxel set. The averaged TAC was analyzed with a WNLLS estimator and the estimated micro-parameters were assigned as the mean of the prior. This procedure was used to mimic a real-world scenario, where the averaged TAC obtained from the cluster or the ROI is used to infer the estimates at the voxel level. The estimation with VB estimator was carried out for each synthetic voxel and several values of λ. To identify the optimal λ, the prior precision was set to the square inverse of $\lambda\,\boldsymbol{m_0}$, according to Eq. 10, with λ varying from 5% up to 200% with steps of 5%. The criteria to choose the optimal λ was the absolute mean

percentage error between micro-parameters and synthetic voxels. A more detailed description of the evaluation is reported in the next section.

2.7. Assessment of simulated data results

For each simulated scenario, both VB and WNLLS estimates were corrected for unreliable estimates following the same criteria used in both simulations.

The comparison was then conducted at the intersection of voxels where both methods gave reliable estimates. The model kinetic rate constants and the macro-parameters of interest were compared with the correspondent true values. Percentage mean bias ($\%bias$) and root mean square error ($\%RMSE$) were used as indexes of performance:

$$\%bias = \frac{100}{N}\sum_{i=1}^{N}\frac{(p_i - \bar{p})}{\bar{p}} \tag{11}$$

$$\%RMSE = 100\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{(p_i - \bar{p})}{\bar{p}}\right]^2} \tag{12}$$

where N indicates the number of simulations (N=1,000), $\bar{p}$ indicates the true parameter and p_i indicates the *i-th* estimated parameter.

For the selection of the optimal λ, separately for each tracer, cluster and synthetic voxel micro-parameters estimates, the absolute percentage error was computed. This was averaged across voxels and micro-parameters, obtaining for each cluster an index of performance, function of λ:

$$\boldsymbol{abs\ \%error} = \frac{100}{N}\sum_{i=1}^{N}\sum_{i=1}^{P}\frac{|p_i - \bar{p}|}{\bar{p}}$$

We selected the optimal λ as the mean λ between clusters that minimized the mean absolute percentage error. This parameter can quantify in a single performance index the

sensitivity to the prior mean and precision, since for each cluster the analysis was repeated independently re-generating the prior mean and exploring a wide range of possible levels of precision. We calculated for a single optimal λ, considered as the mean of the optimal λ between tracers, the percentage mean bias (Eq. 11) on the micro-parameters for each tracer and for each cluster.

2.8. Application to *in vivo* positron emission tomography data

We applied VB to clinical PET data, considering both reversible and irreversible tracers, described by compartmental models of different complexity: 1) L[1-^{11}C]leucine, marker for regional rates of cerebral protein synthesis (Bishu et al., 2008); 2) [^{11}C]WAY100635, targeting serotonin 5-HT$_{1A}$ receptor (Bose et al., 2011); 3) [^{18}F]FDG in skeletal muscle, marker for the glucose metabolism (Bertoldo et al., 2006). For each PET dataset we considered three subjects as representative test cases.

The choice of these tracers was based on the compartmental models used in the literature to describe their kinetics. The model structure, equations and parameters of interest are reported in Figure 2. In particular, we considered:

- a two-tissue three-rate constant (3K) model (Fig. 2B), which is here used to describe the irreversible kinetics of L[1-^{11}C]leucine. To note that it has the same complexity of the Sokoloff model to describe [^{18}F]FDG kinetics in brain (Sokoloff et al., 1977);
- the classic 2TCM model (Fig. 2A), used to describe the kinetics of [^{11}C]WAY100635;
- the 5K model (Fig. 2C), for the [^{18}F]FDG in skeletal muscle.

As result of this selection we therefore considered:

L[1-^{11}C]leucine: Three male healthy subjects of a previously published study (age 20 to 24) underwent a 90-min dynamic PET scan in a HRRT (CPS Innovations, Knoxville, TN, USA) scanner after a 2-min intravenous infusion of 20 to 30 mCi of L[1-^{11}C]leucine. The criteria for subject inclusion and the procedure for the PET studies are described in detail in (Bishu et al., 2008). Twenty-one regions of interest were derived as described in Veronese et al. (2010).The PET acquisition facility: NIH PET centre, Bethesda, Maryland (USA).

[^{11}C]WAY100635: Three male healthy subjects of a previously published study (35.7 ± 10.5 years old) underwent a 95-min dynamic PET study in an ECAT EXACT3D (Siemens/CTI, Knoxville, TN, USA) scanner after a bolus injection of 301± 12 MBq of [^{11}C]WAY100635. Full details on PET procedures, arterial data extraction and processing are reported in Bose et al. (2011). Forty-six ROIs were derived from the Hammersmith Brain Atlas (Hammers et al., 2003) as described in Bose et al. (2011). PET acquisition facility: Imanet PET centre, London (UK).

[^{18}F]FDG: Three male healthy subjects of a previous study (36 ± 5 years old, body mass index = 22.6 ± 0.8 kg/m^2) were considered (Bertoldo et al., 2006). The subjects were studied in the fasting state and underwent a 90-min dynamic PET scan in a ECAT HR+ (Siemens/CTI, Knoxville, TN, USA) after injection of 6 mCi of [^{18}F]FDG. Details on subject inclusions, PET procedures and processing are reported in Bertoldo et al. (2006). The ROIs included in the analysis were the anterior tibialis and the soleus muscles, as described in Bertoldo et al. (2006). PET acquisition facility: Pittsburgh PET centre, Pittsburgh (PA, USA).

We want to highlight that all the datasets were acquired independently and therefore experimental settings changed across studies. For example, the brain segmentation was inconsistent between datasets, since the regions definition was selected to best match the

particular tracer tissue distribution. These characteristics represented the best conditions to test the flexibility and robustness of VB.

For all the datasets, voxel-wise estimates of V_T (for 2TCM) and K_i (for 3K and 5K models) were obtained with WNLLS and VB applied to the corresponding compartmental models as in the simulation studies, i.e. including the extended error model.

The priors for VB estimator were generated using the hierarchical approach described above: WNLLS was applied to the ROI TACs and then $\boldsymbol{m_0}$ was set equal to the region-wise WNLLS estimates and $\boldsymbol{\Lambda_0}$ was defined as in Eq. 10, with λ determined from the simulation results.

VB and WNLLS voxel-wise estimates were compared at the intersection of voxels where both methods gave reliable estimates, after correction for unreliable estimates. WNLLS estimates were considered as reference values for the comparison despite its sensitivity to initial estimates and non-convergence in a significant percentage of voxels. Nevertheless, when WNLLS converges, its results are characterized by the same properties as the estimator itself, i.e. non polarization, consistency, asymptotic normality, and efficiency (Cobelli and Carson, 2001).

We compared unreliable estimates percentage, correlation (as Pearson's correlation coefficient R^2), slope and intercept of the regression analysis and the mean relative difference (MRD) between VB and WNLLS estimates as performance indexes. We considered both micro- and macro-parameters of interest. Computational time was measured for each voxel of each subjects of the three datasets available for both VB and WNLLS. Both algorithms were implemented with no parallelism using MATLAB Release 2015b, The MathWorks, Inc., Natick, Massachusetts, United States. The analysis was carried out on a quad core Intel Xeon E5450 Processor (3.00 GHz).

3. Results

3.1. Simulation study 1: performance of VB and WNLLS

Results of the first simulation showed comparable performance in term of bias in the parameter estimates between VB and WNLLS (Table 2). The macro-parameters of interest (V_T for 2TCM and K_i for 5K model) were estimated with negligible bias in the three considered cases, while micro-parameters exhibited slightly higher biases. The micro-parameter V_b in the 2TCM simulation with slow kinetics showed the highest bias for both VB and WNLLS (-10%±27% and -12%±23% respectively). It is interesting to note that for the most complex 5K all the parameters were estimated with a bias smaller than 10%.

Table 2. Percentage bias of VB and WNLLS estimators in 3 different simulated cases (2TCM fast and slow kinetics and 5K kinetics). Bias is reported for each micro- and macro-parameter as mean (SD). The macro-parameters of interest are V_T [ml/cm^3] volume of distribution for the 2TCM and K_i [ml/cm^3/min] irreversible uptake of the tracer in the tissues for 5K model

		K_1 [ml/cm^3/min]		k_2 [1/min]		k_3 [1/min]		k_4 [1/min]		k_5 [1/min]		V_B [unitless]		Macro-parameter	
2TCM fast kinetics	WNLLS	-1	(2)	-1	(5)	3	(4)	3	(2)	-	-	-5	(12)	-1	(1)
	VB	1	(6)	4	(13)	6	(9)	4	(5)	-	-	-3	(10)	-2	(2)
2TCM slow kinetics	WNLLS	1	(7)	-1	(20)	-9	(32)	3	(14)	-	-	-12	(23)	-1	(3)
	VB	0	(9)	-2	(25)	-4	(42)	1	(22)	-	-	-10	(27)	0	(3)
5K kinetics	WNLLS	-2	(5)	-9	(19)	-8	(15)	-7	(19)	-2	(12)	4	(11)	0	(1)
	VB	0	(8)	0	(20)	6	(15)	9	(23)	3	(17)	-6	(19)	0	(2)

When considering the RMSE (Table 3), the difference between VB and WNLLS performance became evident: for both micro- and macro-parameters in all the three considered cases VB yielded lower mean RMSE with significantly lower variability.

Table 3. Root mean square error (RMSE) of VB and WNLLS estimators in three different simulated cases (2TCM fast and slow kinetics and 5K kinetics). RMSE is reported for each micro- and macro-parameter as mean (SD). The macro-parameters of interest are V_T [ml/cm^3] volume of distribution for the 2TCM and K_i [ml/cm^3/min] irreversible uptake of the tracer in the tissues for 5K model.

		K_1 [ml/cm^3/min]		k_2 [1/min]		k_3 [1/min]		k_4 [1/min]		k_5 [1/min]		V_B [unitless]		Macro-parameter	
2TCM fast kinetics	WNLLS	30	(8)	63	(16)	47	(13)	32	(9)	-	-	38	(16)	19	(8)
	VB	20	(4)	38	(9)	33	(8)	26	(5)	-	-	32	(11)	15	(5)
2TCM slow kinetics	WNLLS	50	(122)	131	(407)	132	(178)	94	(170)	-	-	85	(164)	23	(19)
	VB	16	(3)	35	(10)	52	(17)	34	(7)	-	-	34	(7)	11	(3)
5K kinetics	WNLLS	45	(12)	95	(21)	61	(11)	95	(28)	63	(19)	84	(30)	10	(3)
	VB	19	(3)	38	(10)	35	(10)	57	(34)	38	(10)	48	(13)	7	(2)

Also the percentage of outliers was significantly lower for VB than for WNLLS (3.4% vs. 8.5% in the 2TCM with fast kinetics; 8.9% vs. 39.8% in the 2TCM with slow kinetics and 7.9% vs. 14.4% in the 5K kinetics).

To summarize, both VB and WNLLS showed low bias but VB was more accurate (lower RMSE) and had better convergence properties due to the regularization introduced by the priors.

3.2. Simulation study 2: sensitivity analysis

The sensitivity analysis of VB estimator was conducted in a simulated "real-world" scenario. The simulation can be employed as a method to evaluate the optimal precision level of the prior (dependent on λ) on several clusters composed by synthetic voxels. The prior mean was re-generated for each cluster from the averaged TAC obtained from the simulated synthetic noisy voxel TACs. The mean absolute percentage error across micro-parameters and synthetic voxels is reported, as a function of lambda in Figure 3. The analysis was conducted for each tracer kinetics and each cluster independently and showed a similar behaviour. The minimum of the curve was located in a very narrow window, spanning values of λ from 0.4 up to 0.6. A boxplot representing the optimal λ of all simulated clusters, for each tracer kinetics is reported in the bottom panel of Figure 3 and shows a homogeneous optimal λ distribution. The Fast 2TCM simulated dataset provided an optimal λ (across the simulated clusters) of 0.58±0.07, while the Slow 2TCM 0.5±0.05 and the optimal λ for the 5K Irreversible kinetics simulated dataset was 0.48±0.07. Considering all datasets as a whole the optimal value was 0.5±0.07.

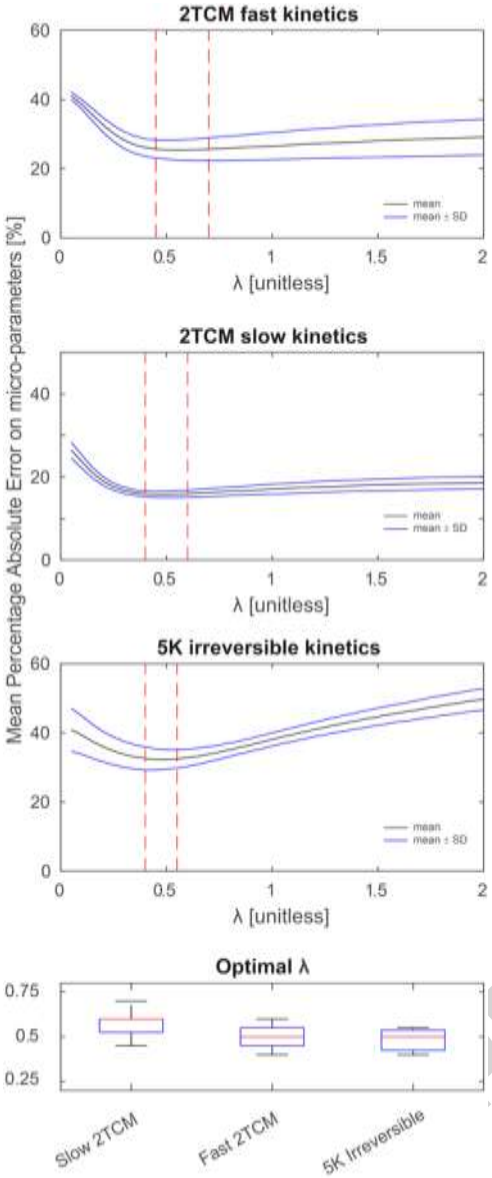

Figure 3. Sensitivity analysis. The mean absolute percentage error is reported in separate graphs, showing respectively the mean (black solid line) and mean ± standard deviation (blue solid line) across clusters as a function of λ for A) 2TCM for tracers with slow kinetics, B) 2TCM for tracers with fast kinetics and C) 5K irreversible kinetics. Optimal λ ranges across clusters are reported with dashed red lines. The bottom panel (D) reports the boxplot of optimal λ for each tracer kinetics across the simulated clusters.

Therefore, based on this empirical approach, the optimal λ suggested to be used in clinical settings with these tracers is 50%, which means that the standard deviation of the priors should be set to half the estimates obtained from the average TACs.

To further address the performances of VB when used in real case settings, we report the percentage bias obtained when the proposed value of λ is used. The bias on the micro-parameters range is comprised in -4.6% and 1.6% for the simulated 2TCM slow kinetics dataset; while for the simulated 2TCM fast kinetics is between -9.5% and 1.7% and lastly the 5K irreversible kinetics dataset has a bias range between -8.8% and 14%.

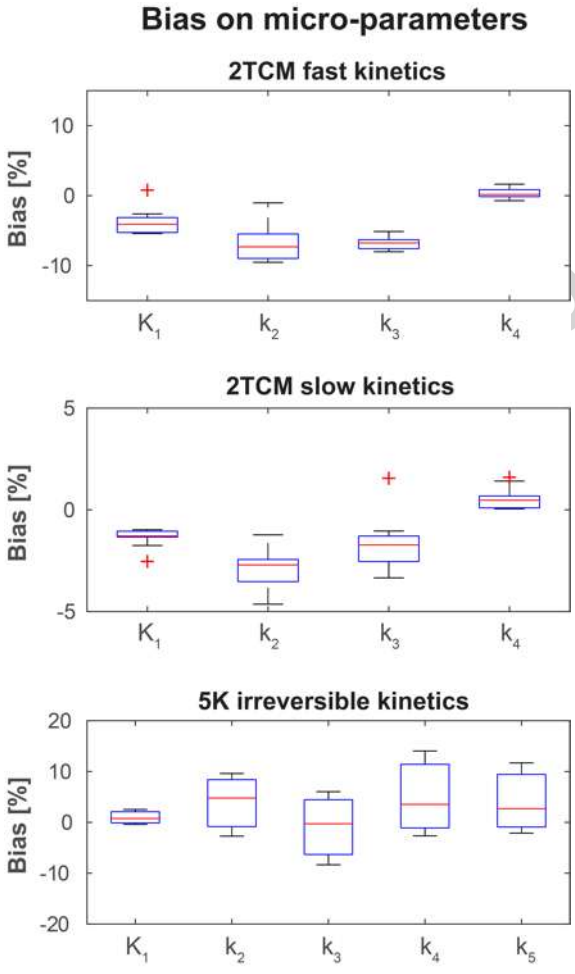

Figure 4. Bias on micro-parameters. The illustration provides a boxplot of the percentage bias on micro-parameters for the simulated tracer kinetics (A) 2TCM fast kinetics (B) 2TCM slow kinetics and (C) 5K irreversible kinetics. The value of $\lambda = 0.5$ was used to provide an insight into the performance of VB in a real scenario when a regional TAC is used to infer the voxel level estimates with the proposed optimal λ value. The values reported in the boxplot are the mean percentage bias calculated across the synthetics voxels of the simulated clusters.

3.3. Application to in vivo positron emission tomography data

Based on the results of the second simulation, we used $\lambda = 0.5$ as measure of variability to set the inverse of the prior variance for all the tracers considered.

The VB results on clinical data confirmed what we previously observed in simulations: the main impact of the method was the percentage of outliers. This can be easily appreciated from the parametric maps of Figure 5, where it is clear that VB allows recovery of the physiological information in the majority of the voxels. In L-[1-^{11}C]Leucine dataset, WNLLS yielded a considerable number of unreliable estimates, i.e. 34%±20%, while on average VB failed to converge to a reliable solution only on 1%±1% of the voxels (Figure 5A). Also for [^{11}C]WAY100635 the number of unreliable estimates was significantly lower for VB than for WNLLS (8%±9% vs 60%±12% respectively) (Figure 5B). The most striking result was obtained with [^{18}F]FDG dataset where WNLLS yielded a percentage of unreliable estimates of 66%±12% while VB only 11%±5% (Figure 5C).

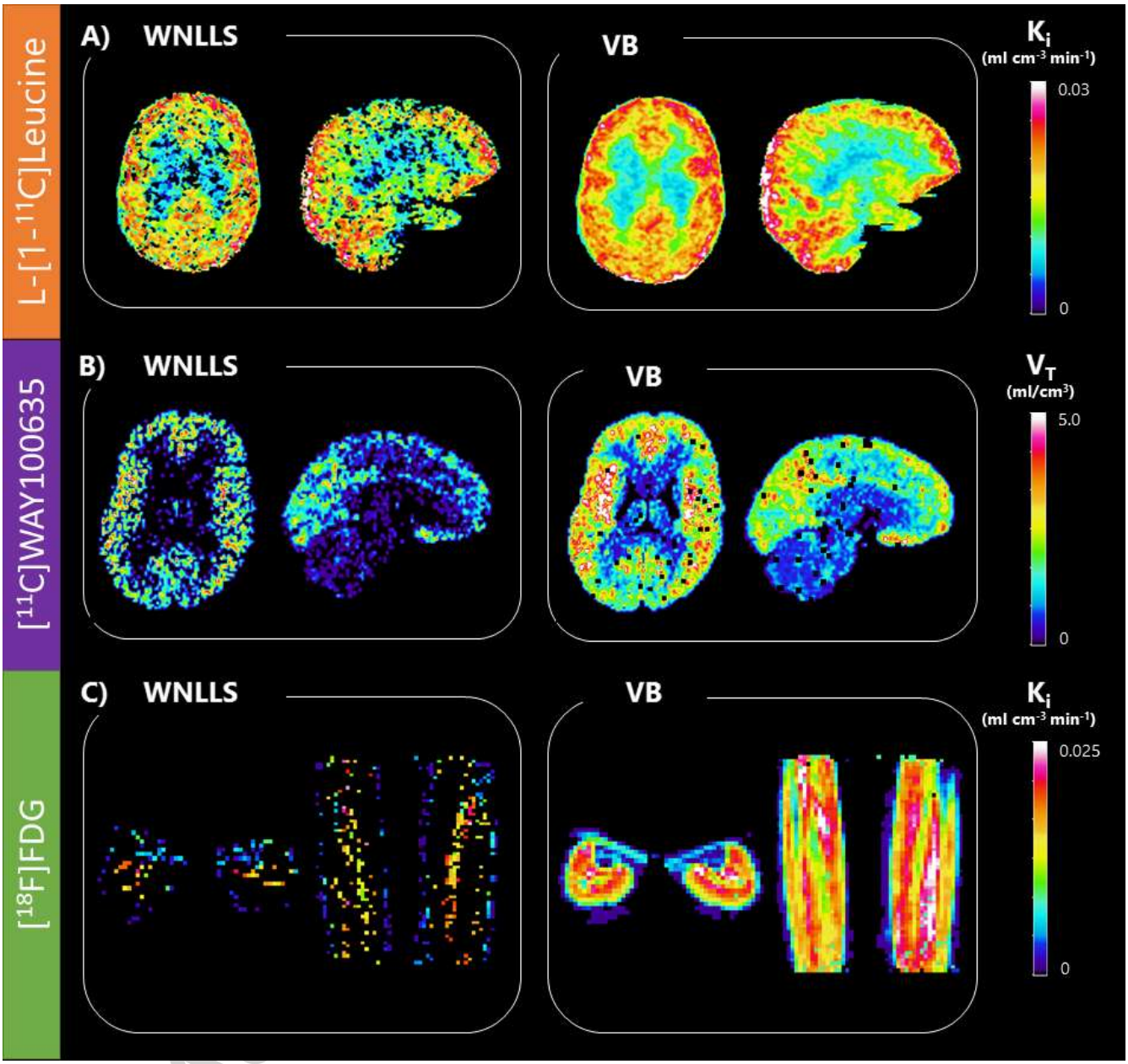

Figure 1. Parametric maps. *The parametric maps of K_i (A, L-[1-^{11}C]Leucine and C, [^{18}F]FDG) and V_T (B, [^{11}C]WAY100635) obtained with WNLLS and VB. For each dataset results refer to a transaxial and a sagittal slice for a representative subject. Parametric maps are presented in their raw form without any smoothing.*

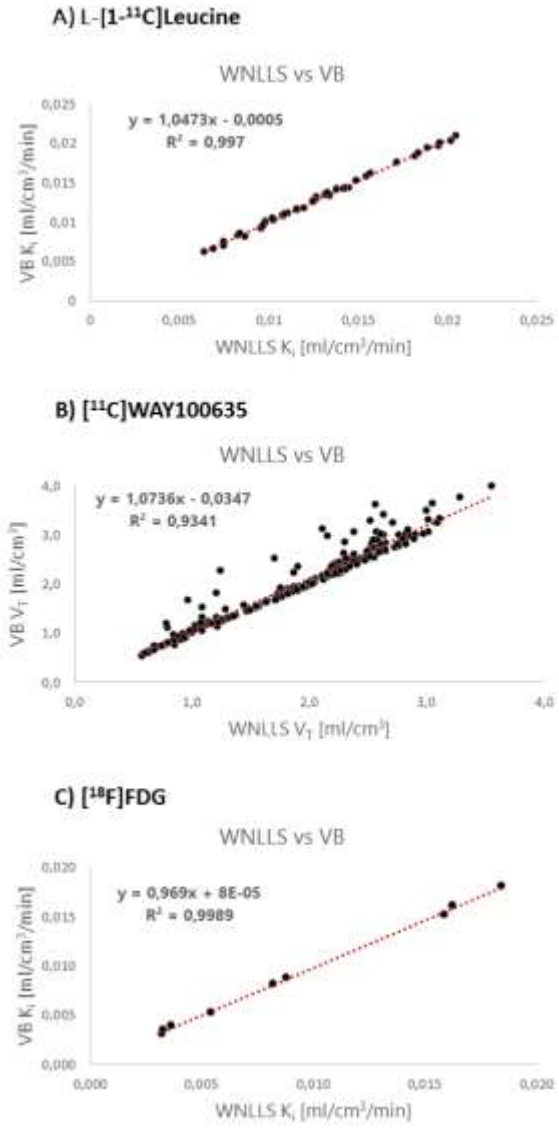

Figure 6. Scatter analysis of K_i and V_T estimates. *The average value of K_i (for L-[1-^{11}C]Leucine (21 ROIs) panel A, and [^{18}F]FDG (3 ROIs), panel C) and V_T (for [^{11}C]WAY100635 (46 ROIs), panel B) obtained within each ROI estimated using WNLLS (x axis) and VB (y axis) for all subjects. Pearson's value R^2, slope and intercept of the fitted regression line are reported.*

The results obtained with WNLLS and VB (where both method converged to reliable solution) were all highly correlated: as regards the L-[1-^{11}C]Leucine dataset, the estimates of K_i obtained with WNLLS and VB showed high correlation and limited mean relative difference (R^2=0.98; MRD=3%±6%; m=1.05; q=-0.0001) (Figure 6A).

Similar results were obtained for the V_T estimates of [^{11}C]WAY100635: these were highly correlated (R^2=0.93) and showed low MRD (5%±12%) between VB and WNLLS (Figure 6B). The regression parameters were $m = 1.07$ and $q = -0.03$.

Also [^{18}F]FDG K_I estimates showed excellent agreement between the two methods (R^2=0.99; MRD=-2%±4%; m=0.97 q=0.0001) (Figure 6C).

Kolmogorov–Smirnov test with a 5% significance level was used to check if weighted residuals produced by VB came from a standard normal distribution. All datasets considered, no evidence of deviation from the model error assumptions was found in the vast majority of voxels (96.3%)**.**

The average calculation time of VB varied from 0.02s to 0.2s per voxel, depending on the model complexity (0.02s per voxel for the 3K model of L-[1-^{11}C]leucine, 0.025s per voxel for the 2TCM of [^{11}C]WAY100635 and 0.2s per voxel for the 5K model of [^{18}F]FDG skeletal muscle). VB was 27%±20% faster than WNLLS for [^{11}C]WAY100635 and 13%±34% faster for [^{18}F]FDG. As regard L-[1-^{11}C]Leucine, computational time was comparable. The bottom line is that the computational time improvement was proportional to both model complexity and noise level. Furthermore, the total time required to complete a scan depended on the number of voxels in the image and the number of samples of the interpolated input function. In fact, the plasma input functions were metabolite-corrected when necessary (Tonietto et al., 2015b) and optimally fitted as in (Tonietto et al., 2015a). The number of samples used for the modelled input function varied across the datasets (from 400 to 800 samples per subject), impacting on the VB computational time. On average VB took 80 min to complete a [^{11}C]WAY100635 brain scan (100'000 voxels), 3.5 hours for a L-[1-^{11}C]leucine HRRT brain scan (650'000 voxels) and 8 hours for a [^{18}F]FDG skeletal muscle scan (150'000 voxels).

4. Discussion

In this work we presented a Variational Bayesian approach for the voxel-wise full kinetic quantification of PET data. The method uses an analytical mean field approximation of the posterior distribution, originally presented by (Chappell et al., 2009) and successfully applied to magnetic resonance imaging and ultrasound data (Rizzo et al., 2016). Here we applied it for the first time to a paradigmatic set of PET tracers considering both simulated and real data.

4.1. Method applicability and performances

We focused our study on brain imaging data, but we extended our analysis to whole body imaging in order to prove the general applicability of the method for PET quantitative parametric mapping.

Indeed, Variational Bayesian approach as set up in this work can be used to solve any compartmental model at the voxel level, and therefore it has wide applicability. We demonstrated this by applying VB to three different models: 2TCM, 3K and 5K models. Since the majority of brain PET kinetic studies employ a 2TCM model for quantification of *in vivo* receptor binding, we considered 2TCM in first instance, by validating the use of VB in both synthetic and real PET data, characterized by different transport rate constants. Then, we evaluated VB performance when applied to a classical but less complex model, i.e. the 3K model, which is the model used in this work to describe L-[1-^{11}C]Leucine. It is worth noting that the 3K model is the predominant compartmental model used to describe [^{18}F]FDG in the brain. Last, we considered one of the most complex model used in PET, i.e. the 5K model exploited to quantify the [^{18}F]FDG PET tracer kinetics in skeletal muscle.

In all cases, VB outperformed the conventional WNLLS by providing accurate and robust estimates, with a substantial lower percentage of outliers. VB accuracy was also compared in one of the data-set included ([^{11}C]WAY100635), to a classic Markov Chain Monte Carlo (MCMC) Bayesian estimator (see supplementary material for more details on this comparison). In Figure S2, V_T maps obtained with both VB and MCMC show a very strong correlation (Pearson R^2=0.99). MCMC does not rely on the mean field approximation exploited by VB. Therefore the high agreement reached between VB and MCMC, confirmed also by previous results (Chappell et al., 2009), suggest that this approximation is indeed reasonable in this context.

Since the compartmental model is solved at the voxel level, VB returns also the maps of microparameters K_1, k_2, k_3, k_4, V_b (and k_5 in the case of [^{18}F]FDG in skeletal muscle). These provides more detailed information about the physiology of the system under study and, furthermore, they allows the calculation of other parameters of interest in addition to the standard V_T or K_i. For example, it is possible to derive the fraction of unlabeled leucine in the precursor pool for protein synthesis $rCPS = (K_i K_1) C_p' / (K_1 - K_i)$ where C_p' is the arterial plasma concentration of unlabeled leucine. In supplementary material, Figure S1 reports the $rCPS$ (L-[1-^{11}C]leucine, panel A), and k_3 ([^{18}F]FDG, panel B). Compared to WNNLS, VB allows the recovery of the physiological information in almost the totality of the voxels.

4.2. Extended noise model

Differently from magnetic resonance imaging, where the noise variance is uniform across the time of the experiment, in PET the non-uniform sampling grid and the radioactive decay of the tracer entail that the level of noise will be different at each time point. Therefore,

we modified the original formulation of VB to adapt it to a non-uniform noise distribution. Analysis of the weighted residuals confirmed the correctness of the noise model.

4.3. Prior definition

Differently from the original work (Chappell et al., 2009), where the priors were set based on typical parameter values from the literature and expected physiological variation, we implemented a data-driven solution for the definition of the prior that are gathered directly from the data. Following the same approach previously published in (Rizzo et al., 2012), we implemented a hierarchical scheme, where the estimates obtained at the region level with a WNLLS estimator, considered the gold standard for PET quantification at high signal-to-noise ratio, were used to define the priors for the voxel level analysis. The region segmentation can be generated either by anatomical atlas segmentation or by functional segmentation using unsupervised clustering.

Another important aspect of the method is the definition of the prior precision $\boldsymbol{\Lambda_0}$**.** We defined $\boldsymbol{\Lambda_0}$ as the square inverse of $\lambda \cdot \boldsymbol{m_0}$, i.e. the prior mean times λ. It is worth noting that the co-variance matrix ($\boldsymbol{\Lambda_0^{-1}}$) used is diagonal, hence that the prior distributions are independent. This, however, will not preclude the posterior probability from being dependent and hence, when supported from the data, the method could account for correlation between rate-constants. Moreover, when prior information on the dependencies between rate-constants is available it would be possible to account for it in the covariance matrix.

The value of λ was set via simulations since it is not possible to infer it directly from the data. In the simulated scenarios we found an optimal value of λ=0.5 (2TCM slow kinetics), λ=0.58 (2TCM fast kinetics) and λ=0.52 ([18F]FDG kinetics), that support the general use of a

value of λ=0.5., which is the value that we suggest to be used also for the quantification of other PET tracers than those presented in the current study.

When applied to the real data in this study, the prior variance defined with λ=0.5 did not constrain the VB estimates to the prior mean value, but on the contrary allowed to detect heterogeneity across the regions of interest (Fig. 7). Also, the percentage differences between the VB estimates and prior mean values were not randomly distributed in the brain, but presented a spatial patterns in agreement with the physiology (e.g. positive (negative) differences in the grey (white) matter for L-[1-11C]leucine or [11C]WAY100635).

We also calculated for each tracer an index of heterogeneity within each region as the ratio between the standard deviation and the mean of the VB estimated macro-parameters within the ROI. Heterogeneity levels varied between 15% and 30% in tissues for L-[1-11C]leucine and [18F]FDG (supplementary tables S1 and S3), whereas was slightly higher for [11C]WAY100635 images, varying from 29% to 69%. However, the [11C]WAY100635 images were segmented with an anatomical atlas of 73 regions, comprising both WM and GM tissues, that can justify the higher heterogeneity levels. The mean value of heterogeneity in all regions was 45% (with a standard deviation of 10%).

We also tested whether the prior variance was too restrictive, therefore artificially reducing the posterior estimates heterogeneity. We repeated the quantification analysis on the real data but doubling the value of λ (from the optimal value of 0.5 found in simulation, to λ =1). In this way, the resulting prior variance was 4-fold the one set originally. We saw a modest increase (maximum 4.6%) of heterogeneity within the ROIs (calculated as the ratio between standard deviation and mean of the VB macro-parameter estimates within the region) for the L-[1-11C]leucine data (Supplementary Figure S3-S4, Supplementary table S1).

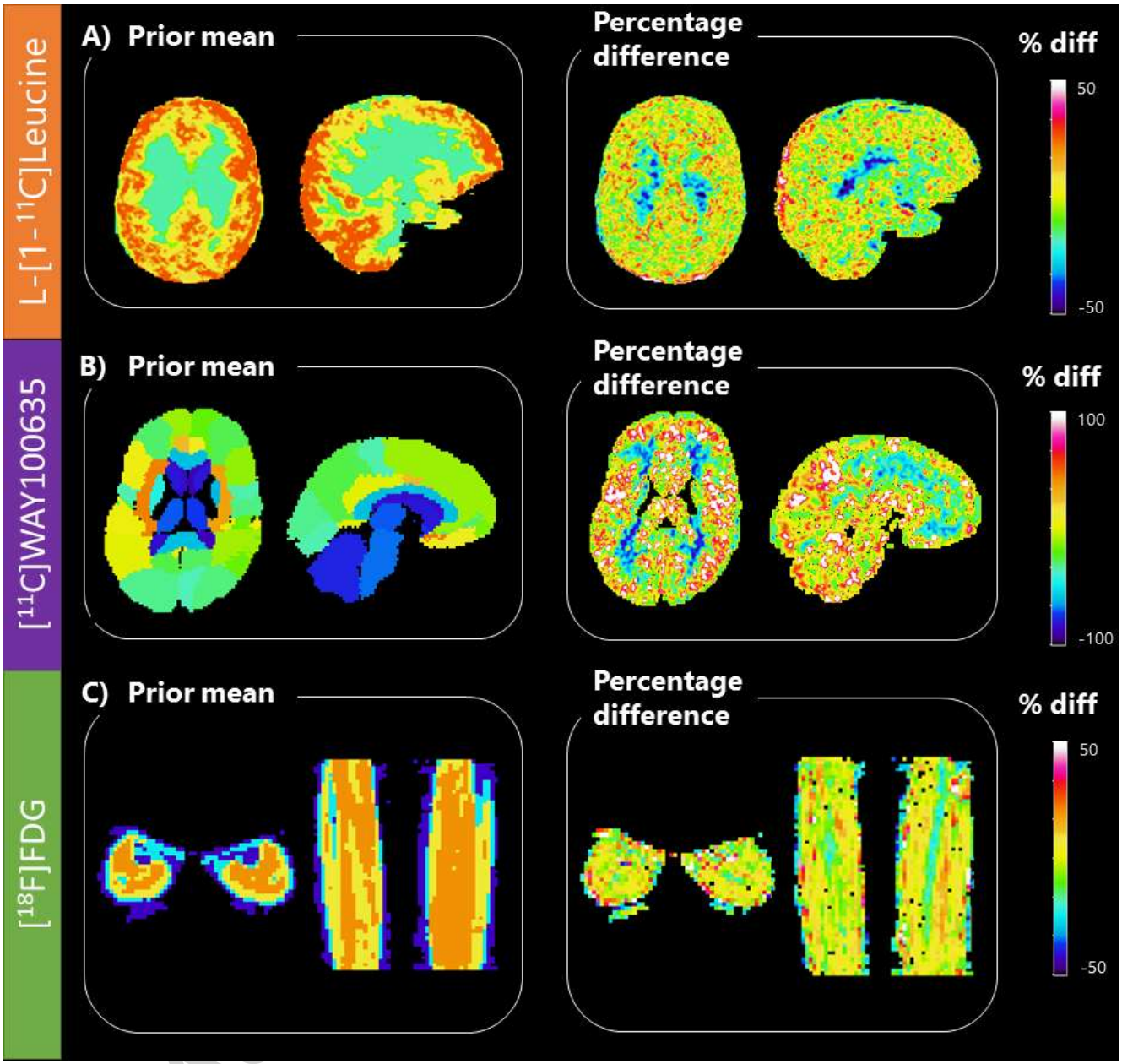

Figure 7. Prior maps and tissue variability. *First Column) Prior mean map obtained for K_i (A, L-[1-^{11}C]Leucine and C, [^{18}F]FDG, obtained after k-means clustering) and V_T (B, [^{11}C]WAY100635, obtained with Atlas Segmentation). Second Column) Percentage differences between the prior mean and the VB estimates (the same reported in Fig. 3) that represents the tissue physiological heterogeneity that VB was able to detect. For each dataset results refer to a transaxial and a sagittal slice for a representative subject. Maps are presented in their raw form without any smoothing.*

Based on these results, we suggest λ=0.5 to be used also for the quantification of other PET tracers than those presented in the current study.

5. Conclusions

Variational Bayesian approach is applied for the first time to PET data. It provides robust and accurate parameter estimates with low percentage of outliers. Computational time required for a whole brain analysis is compatible with clinical practice, even when complex compartmental models are employed.

Acknowledgements

This work was supported by the Padova University grant “Neuroimaging Genetics: Models and Methods to Integrate Brain Phenotype and Genotype” (2013). We thank Dr. Carolyn Smith at the Section on Neuroadaptation and Protein Metabolism of the National Institute of Mental Health for providing the L-[1-^{11}C]leucine images and Simone Zanoni for his help in many useful discussions. Federico E. Turkheimer and Mattia Veronese are supported by the PET Methodology award by the Medical Research Council UK (G1100809/1).

Disclosure

Nothing to declare.

Appendix A

VB updates and Free Energy formulation for PET data

The extension of the noise model affects the VB update equations which become:

$$\mathbf{\Lambda} = sc\boldsymbol{J}^T\mathbf{\Sigma}_e^{-1}\boldsymbol{J} + \mathbf{\Lambda_0} \tag{A.1}$$

$$\mathbf{\Lambda m}_{\boldsymbol{new}} = sc\boldsymbol{J}^T\mathbf{\Sigma}_e^{-1}(\boldsymbol{k} + \boldsymbol{J}\boldsymbol{m_0}) + \mathbf{\Lambda_0}\boldsymbol{m_0} \tag{A.2}$$

$$\frac{1}{s} = \frac{1}{s_0} + \frac{1}{2}(\boldsymbol{k}^T\mathbf{\Sigma}_e^{-1}\boldsymbol{k}) + \mathrm{Tr}\left(\mathbf{\Lambda^{-1}}\boldsymbol{J}^T\mathbf{\Sigma}_e^{-1}\boldsymbol{J}\right) \tag{A.3}$$

$$\mathrm{c} = \frac{N}{2} + c_0 \tag{A.4}$$

where $\mathbf{k}$ represents the vector of the residuals between the data and the model prediction and $\boldsymbol{J}$ is the Jacobian of the model ($\boldsymbol{J_{i,j}} = \partial g_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta_j}$, i.e. the first-order partial derivatives of the model $g(\boldsymbol{\theta})$ with respect to the parameters $\boldsymbol{\theta}$).

Last, also the Free Energy formulation should be adapted:

$$\begin{aligned} F = &-\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{\Sigma_e}| + \frac{N}{2}\left(\log s + \psi(c)\right) - \frac{sc}{2}\left(\boldsymbol{k}^T\mathbf{\Sigma}_e^{-1}\boldsymbol{k} + \mathrm{Tr}\left(\mathbf{\Lambda^{-1}}(\boldsymbol{J}^T\mathbf{\Sigma}_e^{-1}\boldsymbol{J})\right)\right) \\ &- \frac{1}{2}\log\frac{|\mathbf{\Lambda}|}{|\mathbf{\Lambda_0}|} - \frac{1}{2}\mathrm{Tr}\left(\mathbf{\Lambda_0\Lambda^{-1}}\right) - \frac{1}{2}\left((\mathbf{m}-\mathbf{m_0})^{\mathrm{T}}\mathbf{\Lambda_0}(\mathbf{m}-\mathbf{m_0})\right) + \frac{N_p}{2} \\ &- (c-1)\psi(c) + \log s + c + \log\Gamma(c) - \log\Gamma(c_0) - c_0\log s_0 \\ &+ (c_0-1)\left(\log s + \psi(c)\right) - \frac{sc}{s_0} \end{aligned} \tag{A.5}$$

where N_p is the number of model parameters, $\Gamma(x)$ is the Gamma function and $\psi(x) = \mathrm{d}(\log\Gamma(x))/\mathrm{d}x$ is the digamma function, i.e. the logarithmic derivative of the Gamma function.

References

Alpert, N.M., Yuan, F., 2009. A general method of Bayesian estimation for parametric imaging of the brain. Neuroimage 45, 1183–9. doi:10.1016/j.neuroimage.2008.12.064

Attias, H., 2000. A Variational Bayesian Framework for Graphical Models, in: In Advances in Neural Information Processing Systems 12. pp. 209–215.

Beal, M.J., 2003. Variational algorithms for approximate bayesian inference. PhD Thesis 1–281.

Bertoldo, A., Peltoniemi, P., Oikonen, V., Knuuti, J., Nuutila, P., Cobelli, C., 2001. Kinetic modeling of [(18)F]FDG in skeletal muscle by PET: a four-compartment five-rate-constant model. Am. J. Physiol. Endocrinol. Metab. 281, E524-36.

Bertoldo, A., Pencek, R.R., Azuma, K., Price, J.C., Kelley, C., Cobelli, C., Kelley, D.E., 2006. Interactions between delivery, transport, and phosphorylation of glucose in governing uptake into human skeletal muscle. Diabetes 55, 3028–37. doi:10.2337/db06-0762

Bertoldo, A., Vicini, P., Sambuceti, G., Lammertsma, A.A., Parodi, O., Cobelli, C., 1998. Evaluation of compartmental and spectral analysis models of [18F]FDG kinetics for heart and brain studies with PET. IEEE Trans. Biomed. Eng. 45, 1429–1448. doi:10.1109/10.730437

Bishu, S., Schmidt, K.C., Burlin, T., Channing, M., Conant, S., Huang, T., Liu, Z., Qin, M., Unterman, A., Xia, Z., Zametkin, A., Herscovitch, P., Smith, C.B., 2008. Regional rates of cerebral protein synthesis measured with L-[1-11C]leucine and PET in conscious, young adult men: normal values, variability, and reproducibility. J. Cereb. Blood Flow Metab. 28, 1502–13. doi:10.1038/jcbfm.2008.43

Bose, S.K., Mehta, M.A., Selvaraj, S., Howes, O.D., Hinz, R., Rabiner, E.A., Grasby, P.M., Turkheimer, F.E., Murthy, V., 2011. Presynaptic 5-HT1A is related to 5-HTT receptor density in the human brain. Neuropsychopharmacology 36, 2258–65. doi:10.1038/npp.2011.113

Casella, G., 1985. An Introduction to Empirical Bayes Data Analysis. Am. Stat. 39, 83. doi:10.2307/2682801

Chappell, M. a, Groves, a R., Whitcher, B., Woolrich, M.W., 2009. Variational Bayesian Inference for a Nonlinear Forward Model. IEEE Trans. Signal Process. 57, 223–236. doi:10.1109/TSP.2008.2005752

Cobelli, C., Carson, E., 2001. An Introduction to Modelling Methodology, in: Modeling Methodology for Physiology and Medicine. Elsevier, pp. 1–13. doi:10.1016/B978-012160245-1/50002-7

Cobelli, C., Foster, D., Toffolo, G., 2002. Tracer Kinetics in Biomedical Research. Kluwer Academic Publishers, Boston. doi:10.1007/b112199

Cunningham, V.J., Jones, T., 1993. Spectral analysis of dynamic PET studies. J. Cereb. Blood Flow Metab. 13, 15–23. doi:10.1038/jcbfm.1993.5

DiStefano, J.I., 2015. Dynamic systems biology modeling and simulation, 1st ed. Academic Press.

Goodpaster, B.H., Bertoldo, A., Ng, J.M., Azuma, K., Pencek, R.R., Kelley, C., Price, J.C., Cobelli, C., Kelley, D.E., 2014. Interactions among glucose delivery, transport, and phosphorylation that underlie skeletal muscle insulin resistance in obesity and type 2 Diabetes: studies with dynamic PET imaging. Diabetes 63, 1058–68. doi:10.2337/db13-1249

Gunn, R.N., Lammertsma, A.A., Hume, S.P., Cunningham, V.J., 1997. Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. Neuroimage 6, 279–87. doi:10.1006/nimg.1997.0303

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19, 224–47. doi:10.1002/hbm.10123

Kamasak, M.E., Bouman, C.A., Morris, E.D., Sauer, K., 2005. Direct reconstruction of kinetic parameter images from dynamic PET data. IEEE Trans. Med. Imaging 24, 636–650. doi:10.1109/TMI.2005.845317

Koeppe, R.A., Holden, J.E., Ip, W.R., 1985. Performance comparison of parameter estimation techniques for the quantitation of local cerebral blood flow by dynamic positron computed tomography. J. Cereb. Blood Flow Metab. 5, 224–34. doi:10.1038/jcbfm.1985.29

Kotasidis, F.A., Tsoumpas, C., Rahmim, A., 2014. Advanced kinetic modelling strategies: towards adoption in clinical PET imaging. Clin. Transl. Imaging 2, 219–237. doi:10.1007/s40336-014-0069-8

Loeb, R., Navab, N., Ziegler, S.I., 2015. Direct Parametric Reconstruction Using Anatomical Regularization for Simultaneous PET/MRI Data. IEEE Trans. Med. Imaging 34, 2233–2247. doi:10.1109/TMI.2015.2427777

Logan, J., Fowler, J.S., Volkow, N.D., Wolf, A.P., Dewey, S.L., Schlyer, D.J., MacGregor, R.R., Hitzemann, R., Bendriem, B., Gatley, S.J., 1990. Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects. J. Cereb. Blood Flow Metab. 10, 740–7. doi:10.1038/jcbfm.1990.127

Marcone, A., Garibotto, V., Moresco, R.M., Florea, I., Panzacchi, A., Carpinelli, A., Virta, J.R., Tettamanti, M., Borroni, B., Padovani, A., Bertoldo, A., Herholz, K., Rinne, J.O., Cappa, S.F., Perani, D., 2012. [11C]-MP4A PET cholinergic measurements in amnestic mild cognitive impairment, probable Alzheimer's disease, and dementia with Lewy bodies: a Bayesian method and voxel-based analysis. J. Alzheimers. Dis. 31, 387–99. doi:10.3233/JAD-2012-111748

Mazoyer, B.M., Huesman, R.H., Budinger, T.F., Knittel, B.L., 1986. Dynamic PET data analysis. J. Comput. Assist. Tomogr. 10, 645–53.

Patlak, C.S., Blasberg, R.G., Fenstermacher, J.D., 1983. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. J. Cereb. Blood Flow Metab. 3, 1–7. doi:10.1038/jcbfm.1983.1

Peng, J.-Y., Aston, J. a D., Gunn, R.N., Liou, C.-Y., Ashburner, J., 2008. Dynamic positron emission tomography data-driven analysis using sparse Bayesian learning. IEEE Trans.

Med. Imaging 27, 1356–69. doi:10.1109/TMI.2008.922185

Rapisarda, E., Presotto, L., De Bernardi, E., Gilardi, M.C., Bettinardi, V., 2014. Optimized Bayes variational regularization prior for 3D PET images. Comput. Med. Imaging Graph. 38, 445–57. doi:10.1016/j.compmedimag.2014.05.004

Rizzo, G., Tonietto, M., Castellaro, M., Raffeiner, B., Coran, A., Fiocco, U., Stramare, R., Grisan, E., 2016. Bayesian Quantification of Contrast-Enhanced Ultrasound Images with Adaptive Inclusion of an Irreversible Component. IEEE Trans. Med. Imaging 62, 1–1. doi:10.1109/TMI.2016.2637698

Rizzo, G., Turkheimer, F.E., Bertoldo, A., 2013a. Multi-scale hierarchical approach for parametric mapping: Assessment on multi-compartmental models. Neuroimage 67, 344–353. doi:10.1016/j.neuroimage.2012.11.045

Rizzo, G., Turkheimer, F.E., Keihaninejad, S., Bose, S.K., Hammers, A., Bertoldo, A., 2012. Multi-Scale hierarchical generation of PET parametric maps: Application and testing on a [11C]DPN study. Neuroimage 59, 2485–2493. doi:10.1016/j.neuroimage.2011.08.101

Rizzo, G., Veronese, M., Zanotti-Fregonara, P., Bertoldo, A., 2013b. Voxelwise quantification of [11C](R)-rolipram PET data: a comparison between model-based and data-driven methods. J. Cereb. Blood Flow Metab. 33, 1032–1040. doi:10.1038/jcbfm.2013.43

Sokoloff, L., Reivich, M., Kennedy, C., Rosiers, M.H. Des, Patlak, C.S., Pettigrew, K.D., Sakurada, O., Shinohara, M., 1977. THE [14 C]DEOXYGLUCOSE METHOD FOR THE MEASUREMENT OF LOCAL CEREBRAL GLUCOSE UTILIZATION: THEORY, PROCEDURE, AND NORMAL VALUES IN THE CONSCIOUS AND ANESTHETIZED ALBINO RAT. J. Neurochem. 28, 897–916. doi:10.1111/j.1471-4159.1977.tb10649.x

Tang, J., Kuwabara, H., Wong, D.F., Rahmim, A., 2010. Direct 4D reconstruction of parametric images incorporating anato-functional joint entropy. Phys. Med. Biol. 55, 4261–4272. doi:10.1088/0031-9155/55/15/005

Tomasi, G., Bertoldo, A., Bishu, S., Unterman, A., Smith, C.B., Schmidt, K.C., 2009. Voxel-based estimation of kinetic model parameters of the L-[1-(11)C]leucine PET method for determination of regional rates of cerebral protein synthesis: validation and comparison with region-of-interest-based methods. J. Cereb. Blood Flow Metab. 29, 1317–31. doi:10.1038/jcbfm.2009.52

Tonietto, M., Rizzo, G., Veronese, M., Bertoldo, A., 2015a. Modelling arterial input functions in positron emission tomography dynamic studies, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 2247–2250. doi:10.1109/EMBC.2015.7318839

Tonietto, M., Rizzo, G., Veronese, M., Fujita, M., Zoghbi, S.S., Zanotti-Fregonara, P., Bertoldo, A., 2015b. Plasma radiometabolite correction in dynamic PET studies: Insights on the available modeling approaches. J. Cereb. Blood Flow Metab. 0271678X15610585. doi:10.1177/0271678X15610585

Turkheimer, F.E., Hinz, R., Gunn, R.N., Aston, J.A.D., Gunn, S.R., Cunningham, V.J., 2003. Rank-shaping regularization of exponential spectral analysis for application to functional parametric mapping. Phys. Med. Biol. 48, 3819–41.

Veronese, M., Bertoldo, A., Bishu, S., Unterman, A., Tomasi, G., Smith, C.B., Schmidt, K.C., 2010. A spectral analysis approach for determination of regional rates of cerebral protein synthesis with the L-[1-(11)C]leucine PET method. J. Cereb. Blood Flow Metab. 30, 1460–1476. doi:10.1038/jcbfm.2010.26

Veronese, M., Schmidt, K.C., Smith, C.B., Bertoldo, A., 2012. Use of spectral analysis with iterative filter for voxelwise determination of regional rates of cerebral protein synthesis with L-[1-11C]leucine PET. J. Cereb. Blood Flow Metab. 32, 1073–85. doi:10.1038/jcbfm.2012.27

Xia, Y., Wang, J., Eberl, S., Fulham, M., Feng, D.D., 2011. Brain tissue segmentation in PET-CT images using probabilistic atlas and variational Bayes inference. Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2011, 7969–72. doi:10.1109/IEMBS.2011.6091965

Zanderigo, F., Ogden, R.T., Bertoldo, A., Cobelli, C., Mann, J.J., Parsey, R. V, 2010. Empirical Bayesian estimation in graphical analysis: a voxel-based approach for the determination of the volume of distribution in PET studies. Nucl. Med. Biol. 37, 443–51. doi:10.1016/j.nucmedbio.2010.02.004

Zanoni, S., Castellaro, M., Rizzo, G., Bertoldo, A., 2015. Variational Bayesian inference for quantification of brain PET data at the voxel level. University of Padova.

Zhou, Y., Aston, J.A.D., Johansen, A.M., 2013. Bayesian model comparison for compartmental models with applications in positron emission tomography. J. Appl. Stat. 40, 993–1016. doi:10.1080/02664763.2013.772569